PadhAI: Information Theory and Cross Entropy

One Fourth Labs

Information Content

What is Information content?

1. Consider the Random variable SR which maps to the direction in which the sun rises: East, West, North & South.
   a. Now, we are told that P(SR=East) is 1.
   b. Here, this is almost a blatantly obvious truth, thus we can say that the Information Gained here is very low.
2. Consider another Random variable ST, which maps to whether there is going to be a storm today: Yes, No.
   a. Now, we are told that P(ST=Yes) = 1
   b. Here, the information gained is very high as this is a rather surprising(low probability) event
   c. We can almost say that $Information\ Content \propto Surprise$
   d. Or in other words $Information\ Content \propto \frac{1}{P(X=Surprise)}$
   e. Thus, it can be inferred that the information content is a function of the probability of the event
   f. $IC(P(X=S))$ Where IC is information content
3. Now, consider two separate events
   a. X maps to which cricket team won the match: A, B, C, D
   b. Y maps to the state of a light switch: On, Off
   c. Now we are told that Team B won the match AND the light switch is On
   d. The total Information gained is $IC(X=B \cap Y=On) = IC(X=B) + IC(Y=On)$
4. Combining the points from above, we have
   a. $IC(P(X=S))$ (Information Content is a function of probability)
   b. $IC(P(X \cap Y)) = IC(P(X)) + IC(P(Y))$ (From the previous example)
   c. From probability theory, if P(X) and P(Y) are disjoint, then $(P(X \cap Y)) = P(X) \cdot P(Y)$
   d. Therefore $IC(P(X).P(Y)) = IC(P(X)) + IC(P(Y))$
   e. Therefore we need a family of function that satisfy $f(a.b) = f(a) + f(b)$
   f. The log functions satisfy this $log(a.b) = log(a) + log(b)$
5. Now we can write the IC function as follows
   a. $IC(X=A) = log(\frac{1}{P(X=A)})$
   b. $IC(X=A) = log(1) - log(P(X=A))$
   c. $IC(X=A) = -log_2 P(X=A)$ (All the logs use base 2)